DOI: http://dx.doi.org/10.26483/ijarcs.v8i9.5164

ISSN No. 0976-5697

Volume 8, No. 9, November-December 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

IMPLEMENTATION OF FUZZY BASED TRAFFIC SENTIMENT ANALYSIS

Mr. A. Vijay Karthik
Assistant Programmer,
Muthurangam Govt. Arts College,
Vellore -2,India

Abstract: Intelligence Traffic System (ITS) is an emerging application that can solve the main traffic issues such as traffic injuries and fatalities. Monitoring the traffic in real-time is the best way to ensure the public traffic safety. Traffic sentiment Analysis is the method concerns about the issues of traffic in the particular transportation systems based on the information collected from social media. This can enable to collect the comprehensive information about the current status of public traffic and safety. The proposed system can automatically retrieve the information from Facebook, Twitter or other blogs, and then they extract the potential safety topics like accidents or traffic-jam. Finally, it can classify the output based on Neuro-Fuzzy logic. The final result of the proposed model has ensured the better classification compared to the existing method.

Keywords: Traffic Sentiment Analysis, Web mining, Intelligence Transport System, Fuzzy logic.

I. INTRODUCTION

Intelligent Traffic System can serve the public to be more informative, safer and smarter. But it fails to consider the public opinions or views. Traffic sentiment analysis is obligatory to collect and analyses the comments from the public [1][3]. The tremendous advancement of Internet and social networking groups are the main sources of data mining that helps to inspect the public traffic sentiments from Facebook pages, Twitter and etc., We propose the traffic sentiment analysis techniques to enrich the ITS which has shown in Fig.1.

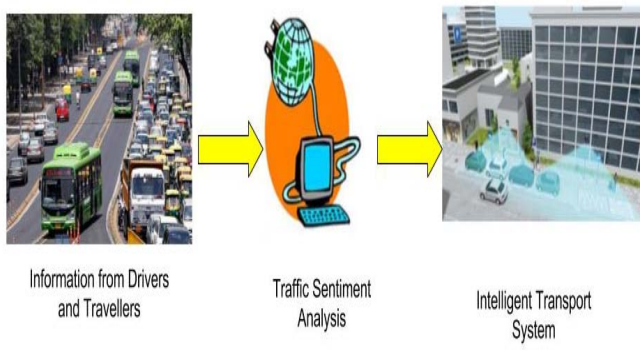

Figure 1. Traffic sentiment analysis based Intelligent Transport System.

TSA is the application of Sentiment Analysis which resolves the traffic issues. In this paper, we also discussed the general problem in TSA to construct the efficient intelligent model. The proposed TSA model deals the issues in different perspective. Fig.1 illustrates the three different stages in ITS. 1) Data Collection: In general many people are free to suggest their opinions in the interactive press or online news on current events. Data collection stage is the process of gathering traffic related information from social media. Adverb- Adjective-Noun (AAN) based technique has been deployed for feature extraction 2) Data sensing: Sensing the retrieved features and analyze the same using Fuzzy based TSA. 3) Decision Making: Based on polarity and Ranking values of features the system can predict some social events.

The paper is arranged as follows: the next section reviews the existing model. Section III covers the problems in TSA. Section IV presents the proposed architecture and algorithm. Section V describes the implementation and Results. Finally, section VI confined with the conclusion.

II. RELATED WORKS

There are number of sentiment analysis approaches discussed in different applications in different levels. Here we mentioned some noteworthy works.

A. Document level

It checks the polarity for the entire document whether it is positive or negative. For example, product reviews in online shopping sites gives the comprehensive view of the product whether good or bad [4]. The main drawback in this method is they provide only polarity result but it fails to extract different features from there.

B. Sentence Level

This method decides the polarity of good, bad, neutral result in each sentence in the document. They classify the sentence in to objective and subjective [2]. Objective indicates the entity information and subjective express the emotional information of particular object. Sentence level sentiment analysis is quite easy, despite it will become complex if a sentence having multiple opinions.

C. Entity level

This is the finer grained approach which solves the problems in existing levels. It is also known as feature-based sentiment analysis. It doesn't consider about the structure of documents. They simply extract the features and find the polarity of each one.

D. Ku's Algorithm

This algorithm can support for both sentence and document level. First, the algorithm detects the sentiment words and then they find the resultant polarity. They tested this in Chinese documents. For that, they created a Chinese seed vocabulary Dictionary. They also extended this method to learn about sentiment words and their strength. They have collected two different sets of sentiment words 1) General Inquiries (GI), 2) Chinese Network Sentiment Dictionary (CNSD). GI is in English language and it should translate into Chinese. CNSD contains the Chinese sentiment words which are collected from the internet. These word resources have used to create the seed vocabulary.

E. Rule based sentiment Analysis algorithm (R-BSA)

It is based on rule-mining algorithm that has determined the efficient rule to extract the features and opinion for a particular product (ex. Price, design, quality, etc.) All inputs are given by human that are collected from World Wide Web. They have used some seed sentiment words to automatically extract the opinion of a sentence. It provided better results compared to Ku's algorithm.

III. PROBLEMS IN TSA

TSA adopts the sentiment analysis to enrich the ITS space. The two main approaches are 1) learning based approach 2) rule based an approach that identifies the features from the web content. Still, there are some practical limitations to implement the Traffic Sentiment Analysis.

1. Data doesn't have a standard structure they may vary in size.
2. Stylistics features vary with people. In some situations, the same text or word possesses different emotions. So we have to consider that also in our analysis. It is complex to implement in some languages [10][11].

A. Selecting a Template Learning based approach

It does not need any prior knowledge to build the bases simply they use some trained classifier [5], [6]. They require large set of datasets with positive and negative polarity for training phase. It is very costlier to implement. As we discussed earlier, the same word having different emotions, so the training data set has to cover these characteristics.

B. Rule-Based Approach

Rule based approach is independent of document size and sentence length [7]. Basic rules of many languages are relatively static. It can easily extend by update sentiment lexicon, despite it provides low precision if we not consider some context in data [8], [9].

IV. PROPOSED METHOD

We have implemented the proposed TSA model in fuzzy rule based approach. Figure 2 shows the architecture that includes 1) data pre-processing 2) sentiment keyword extraction 3) fuzzy-rule based classification. In addition with that the proposed model includes the Fuzzy based classification in order to improve the precision value.

In data pre-processing, we extract the sentiment words from the document"d". Then calculate weights for each word based on our predefined word base. In some case, the words may contain some adverbs such as "most", "very", "more" and "over". For that we are assigning grades for such adverb. Whenever, the adverb comes with sentiment word its degree has been multiplied by the grade value. The steps of the proposed model have explained below.

Fuzzy-Rule Based sentiment Analysis

Input: Obtain a document d

Do pre-process on d

Do the extraction on d

Count the word $\{w_1, w_2, w_3, \ldots..w_N\}$

For each word w_i

Do the option strength calculation

Do the degree calculation di, $\sum_{i=1}^{n} d_i = 1$

End for

Do fuzzy rule based classification

Output: Sentiment polarity of document d.

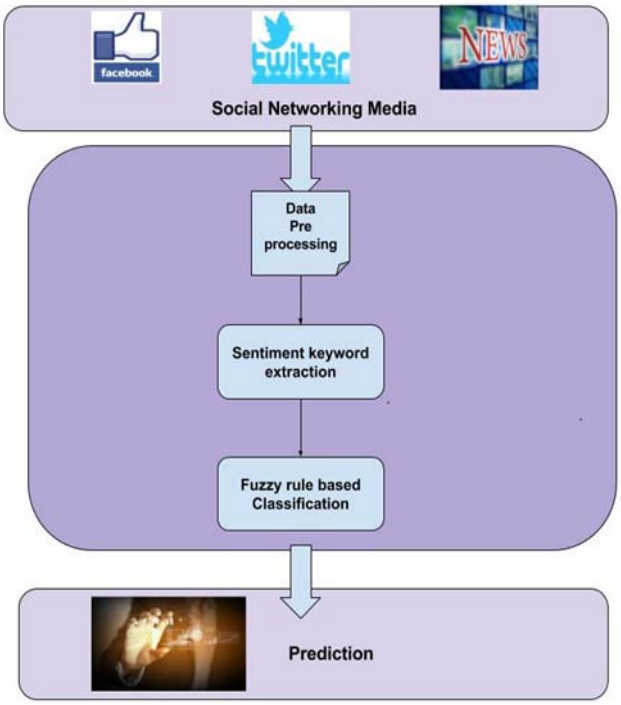

Figure 2. Fuzzy –Rule based TSA Architecture

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. The parameters of fuzzy rule based model

In fuzzy rule classification classifies the data based on five input parameters, such as Mean, maximum value, minimum value, standard deviation, variance.

a) Mean (μ)

Mean is calculated as sum of all data weights x_1, x_2, x_3….x_N in the dataset divided by the total number of data (N).

$$Mean(\mu) = \sum x/N$$

b) Maximum Value

We have to find the maximum value among N datas in the dataset.

$$Max = Maximum(x_1, x_2 \ldots\ldots x_N)$$

c) Minimum Value

We have to find the minimum value among N datas in the dataset.

$$Min = Minimum(x_1, x_2 \ldots\ldots x_N)$$

d) Standard deviation (SD)

Standard deviation is to find the variation from the average value for each data in the dataset. In this method, we can find the data points which are closer to mean value.

$$Std.deviation(SD) = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

e) Variance

It is the square of standard deviation. It measures the variation of each data points in dataset.

$$Variance = \frac{\sum(x-\mu)^2}{N}$$

B. Fuzzy Rule based Model

It is IF-THEN based classification that needs to calculate all parameters such as average, Maximum, Minimum, Standard deviation and variance for each extracted words. Then we classify the word as positive or negative polarity based on fuzzy rule. It is the five inputs and one output model. We defined the threshold value of 0.5 to classify the features in to two different polarities. We have chosen 0.5 threshold because it has only 2 possible outcomes either positive or negative. If the value is greater or equal to 0.5 it is considered as good or else it is bad.

a) If all parameters are bad then the classification is bad.

b) If all parameters are good then the classification is good.

c) If anyone parameter has changed then we consider the maximum values as the classification output.

V. IMPLEMENTATION & RESULT

The fuzzy based TSA model has implemented using JAVA and tested the proposed algorithm in some online datas. Also, we have compared our model with existing TSA model. We proved our classification model has provided better results and precision values than the rule-based TSA model. We have tabulated the results in table1.

Here we compared our fuzzy model results with simple ITS and Rule based TSA models using different parameters as accuracy, error rate and classification time period. From the above, we have proved that our model shows better error rate and accuracy.

Table1: Comparison of proposed output values with existing model

Algorithm	Accuracy	Time period	Error rate	Recall	Ranking
ITS	79.34	5.21	0.053	0.62	89.54
TSA	85.23	4.43	0.047	0.54	93.21
Fuzzy_ TSA	94.2	3.12	0.023	0.34	97.52

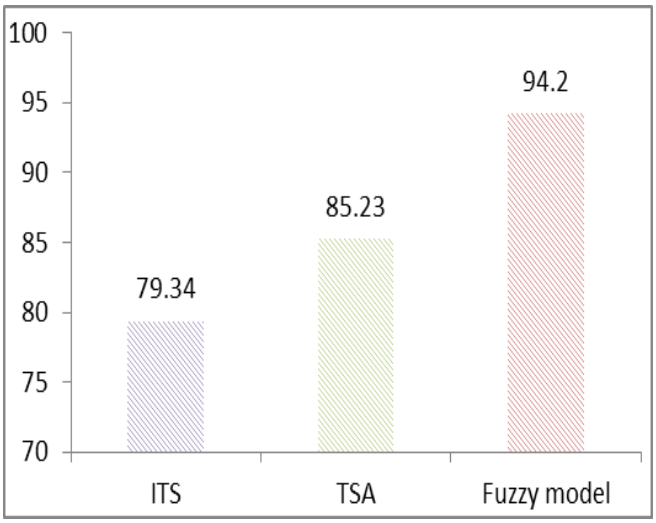

Figure 3. Accuracy comparison

VI. CONCLUSION

Datas from social-media mostly are in unstructured and unorganized manner. So retrieving sentiment information from that is a difficult task. In despite of that, we address our novel fuzzy based TSA model to solve the traffic issues in the Intelligence Transport System. It takes the lesser time for classification and also better error rate. We have tested our algorithm with different kind of datasets, which manifests better results in all. In our future work, we also focus about some unofficial style of text (“tat”,”wot”,”y”) classification.

REFERENCES

[1] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Retrieval, vol. 2, no. 1/2, pp. 1 –135, Jan.2008.

[2] B. B. Khairullah Khan, Aurangzeb Khan, "Sentence based sentiment classification from online customer reviews," ACM, 2010.

[3] Jianping Cao, Ke Zeng, Hui Wang, "Web-Based Traffic Sentiment Analysis: Methods and Applications", IEEE Transactions on Intelligent Transportation Systems, Vol. 15, No. 2, April 2014.

[4] Parneet Kaur, Bindu Rana (2011),"Analysis of Student Physical Health Data using Data Mining Algorithm", Proceeding of International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.6 No.18, pp.2325-2326.

[5] Shitong W., Chung KFL, Hongbin S.(2005),"Fuzzy taxonomy, quantitative database and mining generalized association rules". Proceedings of Intelligent Data Analysis Journal, Vol-9, issue-2, pp 207-217.

[6] Kang Hanhoon, YooSeongJoon, Han Donglil,"Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews" Expert Syst Appl, 39:6000-10, 2012.

[7] P.D.Turney,"Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification reviews,"inProc.40th annu. Assoc. Comput.Linguist., 2002, pp.417-424.

[8] C.Whitelaw, N.Garg, and S.Argamon, "Using appraisal groups for sentiment analysis." In Proc14th ACM Int.Conf.Inf.Knowl.Manage.,2005,pp.625-631.

[9] F. Y. Wang, "Social computing: Concepts, contents, and methods," Int. J. Intell. Control Syst., vol. 9, no. 2, pp. 91–96, 2004.

[10] F.-Y. Wang, R. Lu, and D. Zeng, "Artificial intelligence in China," IEEE Intell. Syst., vol. 23, no. 6, pp. 24–25, Nov./Dec. 2008.

[11] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favourability using natural language processing," in Proc. 2nd Int. Conf. Knowl. Capture, 2003, pp. 70–77.